Manuscript in Press, *Emotion*

# Automatic Vigilance for Negative Words is Categorical and General

Zachary Estes & James S. Adelman

University of Warwick

Word count: 3415

Correspondence to:

Zachary Estes

Department of Psychology

University of Warwick

Coventry CV4 7AL U.K.

Email: z.estes@warwick.ac.uk

Phone: +44 (0) 247 652 3183

Abstract

With other factors controlled, negative words elicit slower lexical decisions and naming than positive words (Estes & Adelman, this issue). Moreover, this marked difference in responding to negative words and to positive words (i.e., between-category discontinuity) was accompanied by relatively uniform responding among negative words (i.e., within-category equivalence), thus suggesting a categorical model of automatic vigilance. Larsen, Mercer, Balota, and Strube (this issue) corroborated our observation that valence predicts lexical decision and word naming latencies. However, on the basis of an interaction between linear arousal and linear valence, they claim that automatic vigilance does not occur among arousing stimuli and they purport to reject the categorical model. Here we show that (1) this interaction is logically irrelevant to whether automatic vigilance is categorical, (2) the linear interaction is statistically consistent with the categorical model, (3) the interaction is not observed within the categorical model, and (4) despite having five fewer parameters, the categorical model predicts word recognition times as well as the interaction model. Thus, automatic vigilance is categorical and generalizes across levels of arousal.

KEYWORDS: arousal, automatic vigilance, categorical processing, lexical decision, valence, word naming

Previously, we demonstrated that the affective factors of arousal and valence predict lexical decision and word naming times (Estes & Adelman, this issue). Responses were faster for highly arousing words than for less arousing words, and for positive words than for negative words. The effect of valence on word recognition was of particular importance, as it provided strong evidence of automatic vigilance (described below). Indeed, by controlling lexical factors such as length and frequency (see Larsen, Mercer, & Balota, 2006) among more than a thousand words, our analysis provided the most powerful demonstration yet of automatic vigilance. Larsen, Mercer, Balota, and Strube (this issue) report an alternative analysis of the same dataset, and their results corroborated our observation that valence predicts lexical decision and word naming latencies. Both analyses therefore indicated that affective factors influence lexical processing, and both analyses supported the validity of automatic vigilance. However, the two analyses diverge critically in the presumed sensitivity and generality of automatic vigilance. Below we evaluate whether automatic vigilance (a) is sensitive to degrees of negativity, and (b) generalizes across levels of arousal. First, though, we review the evidence of automatic vigilance more generally.

## Automatic Vigilance

Stimuli are automatically evaluated as negative (aversive) or positive (appetitive; see Fazio, 2001; Klauer & Musch, 2003). Such stimulus evaluation occurs immediately (Lazarus, 1982; Zajonc, 1980), thereby facilitating rapid avoidance and approach behaviors (Lavender & Hommel, 2007; Neumann, Forster, & Strack, 2003). Negative stimuli in particular are of paramount urgency, as the failure to avoid a negative stimulus, such as a predator, may be fatal. Failure to attain a positive stimulus, such as prey, is less likely to be fatal because additional opportunities may be forthcoming (Pratto & John, 1991). Consequently, following initial stimulus evaluation, attention is disengaged more slowly from negative stimuli than from neutral or positive stimuli (e.g., Fox, Russo, Bowles, & Dutton, 2001; Horstmann, Scharlau, & Ansorge, 2006; McKenna & Sharma, 2004). This prolonged attentional monitoring of negative stimuli, termed *automatic vigilance*, produces slower responses to negative stimuli than to positive stimuli on most cognitive tasks (e.g., Algom, Chajut, & Lev, 2004; Pratto & John, 1991; Wentura, Rothermund, & Bak, 2000).

*The Categorical Model.* Failing to recognize a predator as such can be fatal. Discriminating the precise level of threat posed by that predator, in contrast, is less urgent. A categorical model of automatic vigilance therefore asserts that, in order to minimize the likelihood of catastrophic error, responding tends to vary more between affective categories than within each category (Juslin & Laukka, 2003). Essentially, all negative stimuli are treated as threatening, since the benefit of quickly averting an extremely dangerous stimulus outweighs the cost of overreacting to a mildly threatening stimulus. For instance, it is safer to flee unnecessarily from a hyena than to be caught deliberating whether an approaching lion is truly threatening. By this categorical model, slightly negative and extremely negative stimuli elicit equivalent levels of automatic vigilance.

The hallmark of categorical perception is that a difference between stimuli is perceived to be larger if the two stimuli are from different categories than if they are from the same category. Emotions, whether expressed visually (e.g., Etcoff & Magee, 1992; Young, Rowland, Calder, & Etcoff, 1997) or vocally (Laukka, 2005), exhibit such categorical perception. Consider a stimulus set consisting of a sad facial expression, a happy expression, and several morphed expressions that vary parametrically in their degree of sadness or happiness. An expression that is 60% sad (and 40% happy) is judged more similar to an expression that is 80% sad than to one that is 40% sad (Bimler & Kirkland, 2001). Moreover, that 60% sad expression is also discriminated less accurately from the 80% sad expression than from the 40% sad expression (Etcoff & Magee, 1992; Laukka, 2005; Young et al., 1997).

Thus, affective differences between stimuli are attenuated within a given category and are accentuated between categories.

The categorical model of automatic vigilance thus yields a clear prediction: If stimulus evaluation is categorical, then response times should vary between affective categories but should be relatively constant within each affective category. In considering the evidence of such a model of automatic vigilance, it may be informative to discuss more generally the properties of a categorical model. The primary property is a discontinuity in the response distribution between stimulus classes. Any process that has a categorical component must exhibit such discontinuity. In terms of automatic vigilance, the boundary between negative stimuli and positive stimuli should be sharp, indicated by a marked difference in response times (i.e., a steep slope) between slightly negative (e.g., “needle”) and slightly positive stimuli (e.g., “candy”). Discontinuity, then, corresponds to the accentuation of differences between categories. A secondary property of categorical models is relatively uniform responding within a given stimulus class (i.e., equivalence). Regarding automatic vigilance, extremely negative (e.g., “poison”) and slightly negative stimuli (e.g., “needle”) should elicit responses that are equally slow (i.e., a flat slope), and slightly positive (e.g., “candy”) and extremely positive stimuli (e.g., “passion”) should elicit responses that are equally fast. Equivalence therefore corresponds to the attenuation of differences within categories.

The categorical model of automatic vigilance thus predicts a flat slope running from extremely negative to slightly negative stimuli (i.e., equivalence), followed by a steep decline to slightly positive stimuli (i.e., discontinuity), and finally a flat slope running from slightly positive to extremely positive stimuli (i.e., equivalence). Pratto and John (1991) directly tested this prediction. In two Stroop color naming experiments, they presented words that were extremely, moderately, or slightly negative or positive. Both experiments revealed an automatic vigilance effect that was characterized by discontinuity between categories and equivalence within categories: Responses were slower to negative words than to positive words, but they were no slower for extremely negative words than for moderately or slightly negative words. Results therefore supported the categorical model.

*Estes & Adelman (this issue).* Our analysis corroborated this observation of a categorical relation between stimulus valence and response times. With a much larger and better controlled stimulus set (including arousal, word length, word frequency, orthographic *N*, and contextual diversity as covariates), we found that the relation between valence and word recognition could be modeled as a step function. As shown in Figure 1 of Estes and Adelman (this issue), we observed a marked difference between negative words and positive words (i.e., discontinuity), but relatively little variance among negative words and among positive words (i.e., equivalence). Regression analyses confirmed that valence explained significantly more variance in both word naming and lexical decision latencies when valence was treated as a categorical factor (i.e., negative or positive) than when it was treated as a continuous factor (i.e., degrees of negativity or positivity). Thus, our analysis revealed the between-category discontinuity and within-category equivalence that are the signature of categorical processing.

Because the primary purpose of our study was to demonstrate that valence influences lexical processing times, we simply covaried out all other factors known to correlate with lexical and affective processing. This allowed us to demonstrate that the effect of valence on response times was not attributable to any of those other factors. Thus, *with other factors controlled*, negative words elicit slower lexical decisions and naming than positive words. Moreover, this automatic vigilance does not appear sensitive to degrees of negativity; automatic vigilance appears to be categorical. However, this approach does not address whether automatic vigilance generalizes across those other factors. To do that, one must test whether the effect of valence on response times varies across levels of the other factor(s) of interest. For instance, if valence were shown to have no effect among highly arousing words, this would indicate that automatic vigilance is limited to less arousing words. This is precisely what Larsen and colleagues claim.

*Larsen et al. (this issue).* Larsen and colleagues tested whether automatic vigilance generalizes across levels of arousal. Using the same dataset that we had analyzed (Estes & Adelman, this issue), Larsen and colleagues treated valence as a continuous factor and included five additional nonlinear and interaction factors: (i) squared valence, (ii) cubed valence, (iii) arousal by linear valence interaction, (iv) arousal by squared valence interaction, and (v) arousal by cubed valence interaction. Although they failed to find any interaction of arousal and valence in word naming latencies, an interaction was observed in lexical decision latencies. Specifically, Larsen et al. found that among non-arousing stimuli positive words elicited faster lexical decisions than negative words, but among highly arousing stimuli this automatic vigilance effect “largely disappears” (p. ###). Despite obtaining an interaction in only half of their analyses, Larsen and colleagues nevertheless concluded that automatic vigilance does not generalize across levels of arousal. Furthermore, they also claimed that this interaction signaled a rejection of the categorical model, in that not all negative words elicit slower responding than positive words. In the following we therefore evaluate separately the two components of Larsen et al.’s claim that automatic vigilance is neither (i) categorical nor (ii) general.

## Theoretical Evaluation

To clarify, when arousal is controlled, valence *does* exert a significant effect on response times (Estes & Adelman, this issue; Larsen et al., this issue). Larsen et al. thus do not reject the validity of automatic vigilance. Rather, Larsen and colleagues purport to reject the categorical model of automatic vigilance. Their rejection of the categorical model rests solely on the observation that, when treated as a linear factor, valence interacts with arousal.

Here we explain that Larsen and colleagues' interaction of arousal and valence (1) is logically irrelevant to whether valence has a categorical effect on response times, and (2) is statistically consistent with the categorical model of automatic vigilance. Moreover, we also show that (3) automatic vigilance occurs across levels of arousal, and (4) despite having five fewer parameters, the categorical model predicts response times as well as Larsen et al.'s interaction model. Thus, as explained below, further consideration indicates that automatic vigilance is categorical and generalizes across levels of arousal.

*1.* Larsen et al.'s purported rejection of the categorical model is logically invalid. It does not follow from an interaction of two factors that both factors must be graded (i.e., continuous). Consider for example the well documented sex difference in mental rotation of visual stimuli. This sex difference becomes more pronounced across the lifespan (Voyer, Voyer, & Bryden, 1995), thereby producing an interaction of biological sex and age. However, it does not follow that biological sex is graded, nor that it has a graded influence on mental rotation. Likewise, an interaction of arousal and valence does not in any way suggest that valence has a graded effect on response times. Such an interaction merely indicates that valence exerts a differential effect on response times at different levels of arousal; valence nonetheless could exert a significant categorical effect within each level of arousal. Thus, Larsen and colleagues have conflated sensitivity (i.e., whether automatic vigilance is categorical) with generality (i.e., whether automatic vigilance generalizes across levels of arousal). Larsen et al.'s interaction may be informative of generality, but it is logically irrelevant to the sensitivity of automatic vigilance.

*2.* In a regression analysis, a functional form (usually linear) is imposed on each predictor, and the choice of functional forms determines the underlying model being tested. A significant interaction indicates either that the factors are not additive, or that at least one of the specified functional forms is incorrect. For example, if one assumes a linear form for both arousal and valence, an interaction could indicate either that the two factors are nonadditive or that at least one of the factors is nonlinear. In their analyses Larsen et al. assumed linear forms for both arousal and valence, and from the interaction they concluded that arousal and valence are nonadditive. However, the interaction could simply indicate that arousal and/or valence is nonlinear. Indeed, the interaction between linear arousal and linear valence is perfectly consistent with our prior demonstration that valence has a nonlinear (categorical) effect on response times (Estes & Adelman, this issue).

*3.* To clearly demonstrate that arousal and valence are nonadditive, the interaction of arousal and valence must be shown within the best model (i.e., with the best-fitting functional form). Given our prior demonstration that a categorical valence model significantly outperforms a linear valence model (Estes & Adelman, this issue), an interaction of arousal and valence would only be theoretically informative if it occurred within a categorical valence model. We therefore examined whether arousal interacts with categorical valence in predicting lexical decision times. Valence and arousal ratings were again taken from the ANEW dataset (Bradley & Lang, 1999a) and mean lexical decision times were again taken from the ELP dataset (Balota et al., 2002). Additional predictors were length in letters, length in syllables, word frequency log-transformed (from HAL; Burgess, 1998), word frequency log-transformed and squared (see Balota, Cortese, Sergent-Marshall, Spieler, & Yap, 2004), and orthographic *N* (from ELP). We first conducted a regression with a categorical (binary) form for valence and a simple linear form for arousal. In replication of our previous result (Estes & Adelman, this issue), categorical valence exerted a significant effect on response times [$t(1012) = 5.20$, $p < .0001$], thus confirming the observation of automatic vigilance. Critically, however, categorical valence did not interact with arousal [$t(1012) = 1.64$, $p > .10$].

To be conservative, we also ran several additional regressions with valence treated as a categorical factor and with arousal treated as a quadratic or cubic factor, and as a categorical factor with two, three, four, five or six levels. In each of the analyses categorical valence exerted a significant effect on lexical decision times (all $p < .05$), and in none of these analyses did arousal interact significantly with categorical valence (all $p > .05$). Thus, automatic vigilance generalized across levels of arousal. For illustrative purposes, Figure 1 presents the data from one of these regressions. In this particular analysis we created high, mid, and low arousal groups by splitting the words from ANEW into upper, middle, and lower thirds of arousal ratings. Within each arousal level, negative and positive valence groups were created by treating valence ratings as categorically negative or positive. Figure 1 plots the mean lexical decision latencies for those words, adjusted for the aforementioned control factors of length, frequency, and orthographic *N*. The figure yields three important observations.

*3A.* First, the magnitude of the automatic vigilance effect remained constant across levels of arousal, varying only from 16.1 ms to 17.5 ms. This automatic vigilance effect is evident in Figure 1 as the vertical distance (i.e., along the Y-axis) between the negative mean and the positive mean within each level of arousal. Treating valence as a categorical factor thus reveals that valence produced a constant effect of approximately 17 ms on lexical decision times. That is, the high arousal words exhibited the same automatic vigilance effect as the low arousal words. These data therefore contradict Larsen et al's claim that highly arousing stimuli do not exhibit an automatic vigilance effect.

*3B.* Second, arousal and valence are non-independent (Bradley & Lang, 1999b; Larsen et al., this issue). Highly arousing stimuli tend to be extremely negative or extremely positive, whereas low arousal stimuli tend to be of moderate valence. The non-independence of arousal and valence is evident here in Figure 1 as the horizontal distance (i.e., along the X-axis) between the negative mean and the positive mean within each level of arousal: The high arousal words exhibit the largest distance (i.e., are most extreme) and the low arousal words exhibit the smallest distance (i.e., are least extreme). As explained below, this relationship between arousal and valence has statistical implications.

*3C.* Third, the slope of the automatic vigilance effect varied systematically across levels of arousal. Namely, the slope becomes increasingly steeper from high arousal (-3.56) to mid arousal (-4.74) and to low arousal (-7.41). This difference in slope emerges from the non-independence of arousal and valence, coupled with the constant magnitude of the automatic vigilance effect. That is, because the effect size (i.e., the vertical distance) remained constant while the range (i.e., the horizontal distance) decreased across levels of arousal, the slope consequently increased across levels of arousal. This observation may explain why Larsen and colleagues obtained an interaction: When valence is treated as a continuous factor, these differences in slope produce an interaction of arousal and valence. But the present analyses indicate that when treated properly as a categorical factor, valence does not interact with arousal.

*4.* Above we have shown that the presumed interaction between arousal and valence is logically irrelevant to and statistically consistent with the claim that automatic vigilance is categorical. From a theoretical perspective, then, Larsen et al.'s analyses do not challenge the categorical model of automatic vigilance. From a practical perspective, though, Larsen et al.'s interaction model could nonetheless be preferable if it were shown to predict word recognition latencies better than the categorical model. Thus, to statistically compare the models, we conducted regression analyses with word length (number of letters), word frequency log-transformed (from HAL; Burgess, 1998), word frequency log-transformed and squared (from HAL; Balota et al., 2004), orthographic *N* (from ELP; Balota et al., 2002), arousal (linear, from ANEW; Bradley & Lang, 1999a), and valence (from ANEW) as predictors of z-transformed lexical decision and word naming latencies (from ELP; cf. Larsen et al., this issue). The categorical model, which treated valence as binary, yielded adjusted $R^2$

of 60.28% for lexical decision and 41.69% for naming. The interaction model treated valence as a linear factor, and also included the five additional nonlinear and interaction factors used by Larsen et al. (i.e., squared valence, cubed valence, arousal by linear valence interaction, arousal by squared valence interaction, and arousal by cubed valence interaction). This model yielded adjusted $R^2$ of 60.46% for lexical decision and 41.49% for naming. Thus, despite having five additional parameters, the interaction model failed to explain any additional variance in word recognition times. Or stated conversely, despite its simplicity the categorical model explained as much variance as the interaction model. Parsimony therefore favors the categorical model.

## Conclusions

Below we summarize the evidence pertaining to the sensitivity and generality of automatic vigilance.

*Is automatic vigilance sensitive to degrees of negativity?* Pratto and John (1991) first showed that slightly negative and extremely negative words elicit equally slow responding, and that slightly positive and extremely positive words elicit equally fast responding. This discontinuity between valence categories and equivalence within categories indicates that automatic vigilance is categorical. That is, automatic vigilance is not sensitive to degrees of negativity (or positivity). Our regression analyses (Estes & Adelman, this issue), which included a much larger and better controlled stimulus set, demonstrated unequivocally that a categorical model of valence significantly outperformed a linear model. Larsen et al. (this issue) did not dispute this. Instead, they claimed that an interaction of linear arousal and linear valence—which was observed in lexical decision times but not in word naming—signals a rejection of the categorical model. We have shown that (1) this interaction is logically irrelevant to the categorical model, (2) the linear interaction is statistically consistent with the categorical model, (3) the interaction is not observed within the categorical model, and (4) the interaction model fails to outperform the simpler categorical model. Thus, Larsen and colleagues' analyses do not in any way challenge the categorical model of automatic vigilance. So in summary, although only two studies (i.e., Estes & Adelman, this issue; Pratto & John, 1991) have directly addressed this question, results from both studies suggest that automatic vigilance is insensitive to degrees of negativity.

*Does automatic vigilance generalize across levels of arousal?* Although our original analysis demonstrated an effect of valence on word recognition times when arousal was statistically controlled, it did not test whether this automatic vigilance effect generalizes across levels of arousal. Larsen and colleagues therefore proposed and tested an interaction model of automatic vigilance. From the significant interaction of linear arousal and linear valence they concluded that automatic vigilance does not generalize across levels of arousal. Specifically, they claimed that automatic vigilance occurs among non-arousing words but not among highly arousing words. Here we have shown that the interaction does not occur when the correct functional form is tested, that automatic vigilance was observed among highly arousing words, and that the magnitude of the automatic vigilance effect was constant across levels of arousal (see points 3 and 3A above). Although the non-independence of arousal and valence poses a methodological obstacle (see points 3B and 3C), the present evidence suggests that automatic vigilance generalizes across levels of arousal.

References

Algom, D., Chajut, E., & Lev, S. (2004). A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. *Journal of Experimental Psychology: General*, *133,* 323-338.

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). The English Lexicon Project: A Web-based repository of descriptive and behavioral measures for 40,481 English words and non-words. Available at http://elexicon.wustl.edu.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133,* 283-316.

Bimler, D. & Kirkland, J. (2001). Categorical perception of facial expressions of emotion: Evidence from multidimensional scaling. *Cognition & Emotion, 15,* 633-658.

Bradley, M. M. & Lang, P. J. (1999a). Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.

Bradley, M. M. & Lang, P. J. (1999b). Fearfulness and affective evaluations of pictures. *Motivation and Emotion*, *23,* 1-13.

Burgess, C. (1998). From simple associations to the building blocks of language: Modeling meaning in memory with the HAL model. *Behavior Research Methods, Instruments, & Computers, 30,* 188-198.

Estes, Z. & Adelman, J. S. (this issue). Automatic vigilance for negative words in lexical decision and naming: Comment on Larsen, Mercer, and Balota (2006). *Emotion.*

Etcoff, N. L. & Magee, J. J. (1992). Categorical perception of facial expressions. *Cognition, 44,* 227-240.

Fazio, R. H. (2001). On the automatic activation of associated evaluations: An overview. *Cognition & Emotion, 15,* 115-141.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130,* 681-700.

Horstmann, G., Scharlau, I., & Ansorge, U. (2006). More efficient rejection of happy than of angry face distractors in visual search. *Psychonomic Bulletin & Review, 13,* 1067-1073.

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770-814.

Klauer, K. C. & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation* (pp. 7-49). Mahwah, NJ: Erlbaum.

Larsen, R. J., Mercer, K. A., & Balota, D. A. (2006). Lexical characteristics of words used in emotional Stroop experiments. *Emotion, 6,* 62-72.

Larsen, R. J., Mercer, K. A., Balota, D. A., & Strube, M. J. (this issue). Not all negative words slow down lexical decision and naming speed: Response to Estes and Adelman. *Emotion.*

Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion, 5,* 277-295.

Lavender, T., & Hommel, B. (2007). Affect and action: Towards an event-coding account. *Cognition and Emotion, 21,* 1270-1296.

Lazarus, R. S. (1982). Thoughts on the relationship between emotion and cognition. *American Psychologist, 37,* 1019-1024.

McKenna, F. P., & Sharma, D. (2004). Reversing the emotional Stroop effect reveals that it is not what it seems: The role of fast and slow components. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 382-392.

Neumann, R., Forster, J. & Strack, F. (2003). Motor compatibility: The bidirectional link between behavior and evaluation. In J. Musch & K. C. Klauer (Eds.), *The Psychology of Evaluation* (pp. 371-391). Mahwah, NJ: Erlbaum.

Pratto, F. & John, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality and Social Psychology, 61,* 380-391.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250-270.

Wentura, D., Rothermund, K., & Bak, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality and Social Psychology, 78,* 1024-1037.

Young, A. W., Rowland, D., Calder, A. J., & Etcoff, N. L. (1997). Facial expression megamix: Tests of dimensional and category accounts of emotion recognition. *Cognition, 63,* 271-313.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist, 35,* 151-175.

**Figure 1.** Mean adjusted lexical decision times as a function of arousal (high, mid, low) and valence (negative, positive). Different slopes are observed across levels of arousal, thereby producing the arousal × valence interaction reported by Larsen et al. (this issue). Critically, however, the magnitude of the valence effect (16-18 ms) remains constant across levels of arousal.